Q. J. R. Meteorol. Soc. (2004), **DRAFT**, pp. 1–999 doi: 10.1256/qj.03.n

Semi-Implicit Time-Integrators for a Scalable Spectral Element Atmospheric Model

By FRANCIS X. GIRALDO[1]*
[1] *Naval Research Laboratory, USA*

(Received 10 December 2003; revised 22 November 2004)

SUMMARY

The Naval Research Laboratory's spectral element atmospheric model (NSEAM) for scalable computer architectures is presented. This new dynamical core is based on a high order spectral element (SE) method in space and uses semi-implicit methods in time based on either the traditional second order leapfrog (LF2) or second order backward difference formulas (BDF2). The novelties of NSEAM are: it is the first semi-implicit SE atmospheric model; LF2 or BDF2 are used to construct the semi-implicit method; and the horizontal operators are written, discretized, and solved in three-dimensional Cartesian space. The semi-implicit NSEAM is validated using: five baroclinic test cases; direct comparisons to the explicit version of NSEAM which has been extensively tested and the results previously reported in the literature; and comparisons with operational weather prediction and well-established climate models. A comparison with the U. S. Navy's spectral transform global forecast model illustrates that NSEAM is 60% faster on an IBM SP4 using 96 processors for the current operational resolution of T239 L30. However, NSEAM can accommodate many more processors while continuing to scale efficiently even at higher grid resolutions.

KEYWORDS: Backward Difference Formula Grid Hexahedral Hydrostatic Icosahedral Leapfrog Primitive Equations

1. INTRODUCTION

The current trend in high performance computing has shifted to the development of clustered systems having tens of thousands of processors; the two fastest computers in the world reported at Supercomputing 2004 have 32,000 (IBM) and 10,000 (SGI) processors. Therefore, to fully exploit this type of architecture requires utilizing numerical methods that rely on a decomposition of the global domain into a multitude of smaller subdomains. Methods that rely on domain decomposition are known as local methods whereas those that do not are referred to as global methods because they require the information of the entire global domain in order to operate on a specific subdomain. The best example of a global method is the spectral transform (ST) method. Examples of local methods include the finite difference (FD), finite element (FE), and finite volume (FV) methods. However, the biggest disadvantage of local methods is that they have not been able to compete, in terms of accuracy, with ST methods which have been used traditionally in operational numerical weather prediction (NWP) and climate models.

Spectral element (SE) methods combine the local domain decomposition property of FE methods with the high-order accuracy, and weak numerical dispersion of ST methods. SE methods have shown promise in many areas of the geosciences including: seismic wave propagation (Komatitsch and Tromp 1999), deep Earth flows (Fournier et al. 2004), climate (Thomas et al. 2002), ocean (Iskandarani et al. 2002), and NWP (Giraldo and Rosmond 2004) modeling. These methods are high-order FE methods where the grid points are chosen to be the Legendre-Gauss-Lobatto points. In Giraldo and Rosmond (2004) we introduced a SE atmospheric model with an explicit leapfrog time-integrator that

* Corresponding author: Naval Research Laboratory, Monterey, CA 93943-5502, USA, email: giraldo@nrlmry.navy.mil.

was shown to scale linearly while achieving accuracies similar to those obtained with ST models. However, in that paper it was shown that the explicit SE model would only outperform ST models at resolutions beyond T406. In order to surpass ST models at resolutions below T406 requires upgrading the time-integrator from explicit to semi-implicit. Explicit time-integrators are too inefficient for NWP applications because the fast moving gravity waves require the use of small time steps to maintain stability. In order to ameliorate this rather stringent time step restriction, researchers have discretized the gravity wave terms implicitly in time and the Rossby wave terms explicitly; this is the idea behind the semi-implicit method (Kwizak and Robert 1971).

In this paper we describe a new semi-implicit Eulerian atmospheric model based on: the SE method in space where the horizontal operators are written, discretized, and solved in three-dimensional (3D) Cartesian space; and a discretization in time by the second order leapfrog (LF2) and backward difference formulas (BDF2). Eulerian atmospheric models typically use LF2 for their semi-implicit method; examples include the U. S. Navy (Hogan and Rosmond 1991) and the National Center for Environmental Prediction (Trémolet and Sela 1999). The reasons for experimenting with BDF2 are:

i. BDF2 are absolutely stable in the region of interest for these equations (for real and distinct eigenvalues);
ii. the resulting computational modes are damped and, therefore, no time-filter is required (time-filters diminish the order of accuracy);
iii. the resulting pseudo-Helmholtz matrices have smaller condition numbers than the one obtained with the LF2 method which translates into fewer iterative solves per time step and results in a more efficient model.

However, BDF2 is by no means ideal for this class of equations (Hamiltonians) and their weaknesses will be discussed in Sec. 2.

The advantages of using Cartesian coordinates are: the pole singularity which plagues the equations in spherical coordinates disappears and the numerical model is completely independent from the grid. Because the numerical method is constructed independently from the grid, this then permits any grid to be used including: icosahedral, hexahedral, telescoping, and adaptive unstructured grids. This independence from the grid is not shared by any of the existing and newly proposed global atmospheric models including the FD model in Davies et al. (2004), the FE model in Côté et al. (1998), the FV model in Lin and Rood (1996), the icosahedral models in Randall et al. (2002) and Majewski et al. (2002), the SE model in Thomas et al. (2002), and the ST models in Hack et al. (1992), Hogan and Rosmond (1991), Temperton et al. (2001), and Trémolet and Sela (1999). In fact, the formulations of all these models are restricted to a specific grid geometry.

Element-based Galerkin (EBG) methods, such as the SE method, offer many more benefits in addition to permitting the use of any grid. For example, to switch from quadrilateral to triangular elements merely requires changing the basis functions, and the associated quadrature points and weights as is done in Giraldo and Warburton (2004) which are being considered for future implementation into NSEAM. Changing from globally conservative to locally conservative methods only requires changing the element boundary conditions to account for fluxes as is done in Giraldo et al. (2002); this then simplifies the construction of adaptive solutions in addition to giving a fully conservative method (the development of non-hydrostatic ocean and atmospheric models using this approach is currently underway). In short, whenever a new contribution from approximation theory emerges, the new basis functions (and associated quadrature) can be easily implemented into the existing EBG model; an example is the current work on non-polynomial expansions derived from the prolate spheroidal wave functions (Boyd 2004) which we are currently testing. This flexibility in EBG methods allows an existing model to adapt to the changing needs in science and computing which justifies the further development of SE models and should ensure their longevity.

The objective of the present work is to introduce a new grid point semi-implicit atmospheric model which:

i. is spectrally accurate;
ii. is highly scalable on distributed-memory computers;
iii. allows for the use of any type of grid;
iv. facilitates its continuing augmentation in accuracy and efficiency due to its element-based construction.

The present work essentially extends the explicit time-integrator of NSEAM to semi-implicit. For convergence rates of the discrete horizontal operators the reader is referred to the article by Giraldo and Rosmond (2004).

The remainder of the paper is organized as follows. Section 2 describes the construction of the semi-implicit time-integrators. Section 3 contains: a description of the implementation of the model on distributed-memory computers using the Message-Passing Interface standard; a scalability comparison between NSEAM and the U. S. Navy's Operational Global Atmospheric Prediction System (NOGAPS); and a performance comparison between the LF2 and BDF2 semi-implicit time-integrators of NSEAM. In Sec. 4 the results for the five test cases used to validate the model are presented. Finally, in Sec. 5 we summarize the key findings of this research. For completeness, Appendix A contains a description of the standard semi-implicit method applied to the hydrostatic primitive equations discretized in time by a general second order time-integrator and in space by the SE method in a Cartesian coordinate system.

2. Semi-Implicit Time-Integrators

The governing equations solved in the present work are the hydrostatic primitive equations (HPE). We assume an adiabatic atmosphere (i.e., no diabatic forcing) and thus only take into account the dynamical processes. Let the HPE be written in the following vector form

$$\frac{\partial \boldsymbol{q}}{\partial t} = S(\boldsymbol{q}) \tag{1}$$

where $\boldsymbol{q} = \left(\pi, \boldsymbol{u}^T, \theta\right)^T$ is the state vector containing the prognostic variables, $\boldsymbol{u} = (u, v, w)^T$, T is the transpose operator, and

$$S(\boldsymbol{q}) = -\begin{pmatrix} \boldsymbol{\nabla} \cdot (\pi \boldsymbol{u}) + \frac{\partial}{\partial \sigma}(\pi \dot{\sigma}) \\ \boldsymbol{u} \cdot \boldsymbol{\nabla} \boldsymbol{u} + \dot{\sigma} \frac{\partial \boldsymbol{u}}{\partial \sigma} + \frac{2\Omega z}{a^2}(\boldsymbol{x} \times \boldsymbol{u}) + \boldsymbol{\nabla}\phi + c_p \theta \frac{\partial P}{\partial \pi} \boldsymbol{\nabla}\pi + \mu \boldsymbol{x} \\ \boldsymbol{u} \cdot \boldsymbol{\nabla}\theta + \dot{\sigma}\frac{\partial \theta}{\partial \sigma} \end{pmatrix} \tag{2}$$

is the source vector function. For closure we require the hydrostatic equation

$$\frac{\partial \phi}{\partial P} = -c_p \theta. \tag{3}$$

For completeness, we define the terms contained in these equations. The terms (a, Ω) are the radius and angular rotation of the earth, respectively. The prognostic variables are: $\pi = p_S - p_T$ where p_S is the surface pressure and p_T is the pressure at the top of the model; the wind velocities, $\boldsymbol{u} = (u, v, w)^T$; and the potential temperature, θ. The diagnostic variables are: the vertical velocity, $\dot{\sigma}$; the geopotential height, ϕ; and the pressure, p. Other variables requiring definition are: the Cartesian coordinate of the grid points, $\boldsymbol{x}$; the vertical coordinate, $0 \leq \sigma \leq 1$, defined from the top of the atmosphere to the surface of the planet; the Exner function, P; and the coefficient of specific heat for constant air pressure, c_p . Finally, the term μ is a Lagrange multiplier required only because we use a 3D momentum equation in Cartesian coordinates to represent the corresponding 2D momentum equation in spherical coordinates (see Côté 1988). With the equations defined we can now proceed to the description of the semi-implicit time-integrators.

(*a*) *General Form of Second Order Semi-Implicit Time-Integrators*

Before describing the implementation of the semi-implicit (SI) method it is crucial to understand which terms must be discretized implicitly. The maximum characteristic wave speed of the HPE, Eqs. (1) and (2), is given by $U + \sqrt{\phi}$ where $U = \boldsymbol{u} \cdot \boldsymbol{n}$ is the wind speed along the direction $\boldsymbol{n}$ and $\sqrt{\phi}$ is the speed of the gravity waves. The fastest gravity waves may travel upto six times faster than the fastest wind velocities. In the semi-implicit (SI) method the terms responsible for the propagation of the gravity waves are treated implicitly and the remaining terms explicitly. This essentially slows down the gravity waves which does not adversely affect the medium-range forecast skill because they only carry a small amount of energy. It should be mentioned that there exist alternatives to the SI method with the leading contender perhaps being the Jacobian-free Newton-Krylov method (see Knoll and Keyes 2004, for example) in which the entire set of equations are solved fully implicitly in time. However, in the present work we shall only consider the SI approach.

In Eq. (2) the source terms contributing to the propagation of gravity waves (G) are

$$S^{\mathrm{G}}(\boldsymbol{q}) = -\begin{pmatrix} \boldsymbol{\nabla} \cdot (\pi \boldsymbol{u}) + \frac{\partial}{\partial \sigma}(\pi \dot{\sigma}) \\ \boldsymbol{\nabla} \phi + c_p \theta \frac{\partial P}{\partial \pi} \boldsymbol{\nabla} \pi \\ \dot{\sigma} \frac{\partial \theta}{\partial \sigma} \end{pmatrix}. \tag{4}$$

We then seek a solution to the equations recast in the following form

$$\frac{\partial \boldsymbol{q}}{\partial t} = \left\{ S(\boldsymbol{q}) - \delta L^G(\boldsymbol{q}) \right\} + \delta \left[L^G(\boldsymbol{q}) \right] \tag{5}$$

where the terms inside the curly brackets are time-integrated explicitly, those inside the square brackets implicitly, L^G (defined in Eq. (A.2)) represents the linearization of S^G, and $\delta = 0$ or 1 depending on whether the method is purely explicit or semi-implicit.

To simplify the following discussion, let us write the integration of Eq. (5) in the following general form

$$\boldsymbol{q}^{n+1} = \sum_{m=0}^{1} \alpha_m \boldsymbol{q}^{n-m} + \gamma \Delta t \sum_{m=0}^{2} \beta_m S(\boldsymbol{q})^{n-m} + \delta \gamma \Delta t \sum_{m=-1}^{2} \rho_m L(\boldsymbol{q})^{n-m} \tag{6}$$

where Table 1 lists the associated coefficients corresponding to the BDF2 and LF2 methods where ϑ is the explicit/implicit weighting with $\vartheta = 0.5$ yielding the trapezoidal rule (also known as Crank-Nicholson). The BDF2A method shown in Table 1 is the method proposed by Karniadakis et al. (1991) for the incompressible Navier-Stokes equations and used by Shen and Wang (1999) for the HPE while the BDF2B method was proposed by Hulsten (1996) but has not been used or further studied. Note that unlike the BDF2 methods, the LF2 method requires the application of the following time filter (Asselin 1972)

$$\tilde{\boldsymbol{q}}^n = \boldsymbol{q}^n + \epsilon \left(\boldsymbol{q}^{n+1} - 2\boldsymbol{q}^n + \tilde{\boldsymbol{q}}^{n-1} \right) \tag{7}$$

where $\tilde{\boldsymbol{q}}$ denotes the time-filtered variable with the time-filter weight ϵ.

Method	α_0	α_1	γ	β_0	β_1	β_2	ρ_{-1}	ρ_0	ρ_1	ρ_2
BDF2A	4/3	-1/3	2/3	2	-1	0	1	-2	1	0
BDF2B	4/3	-1/3	2/3	8/3	-7/3	2/3	1	-8/3	7/3	-2/3
LF2	0	1	2	1	0	0	ϑ	-1	$1-\vartheta$	0

TABLE 1. Backward difference formulas (BDF2) and leapfrog (LF2) time-integrators and their associated coefficients corresponding to Eq. (6).

By exploiting the linearity of the L operator we can rewrite Eq. (6) as

$$\boldsymbol{q}^{n+1} = \sum_{m=0}^{1} \alpha_m \boldsymbol{q}^{n-m} + \gamma \Delta t \sum_{m=0}^{2} \beta_m S(\boldsymbol{q})^{n-m} + \delta \gamma \Delta t L \left(\sum_{m=-1}^{2} \rho_m \boldsymbol{q}^{n-m} \right). \tag{8}$$

Furthermore, we can simplify this equation by extracting the fully explicit solution from Eq. (8) as follows

$$\boldsymbol{q}^{\text{explicit}} = \sum_{m=0}^{1} \alpha_m \boldsymbol{q}^{n-m} + \gamma \Delta t \sum_{m=0}^{2} \beta_m S(\boldsymbol{q})^{n-m}. \tag{9}$$

Multiplying Eq. 8 by ρ_{-1}, and adding $\sum_{m=0}^{2} \rho_m \boldsymbol{q}^{n-m}$ yields

$$\boldsymbol{q}_{tt} = \widehat{\boldsymbol{q}} + \delta \gamma \Delta t \rho_{-1} L \left(\boldsymbol{q}_{tt} \right) \tag{10}$$

where

$$\widehat{\boldsymbol{q}} = \rho_{-1} \boldsymbol{q}^{\text{explicit}} + \sum_{m=0}^{2} \rho_m \boldsymbol{q}^{n-m} \tag{11}$$

and

$$\boldsymbol{q}_{tt} = \sum_{m=-1}^{2} \rho_m \boldsymbol{q}^{n-m} \equiv \rho_{-1} \boldsymbol{q}^{n+1} + \sum_{m=0}^{2} \rho_m \boldsymbol{q}^{n-m}. \tag{12}$$

Equation (10) is the form that we use for the construction of the semi-implicit method along with the definitions in Eqs. (9), (11), and (12). The subscript tt in the semi-implicit state vector $\boldsymbol{q}$ is meant to emphasize the similarity between the semi-implicit correction and a temporal second order derivative. This is quite evident for BDF2A where $\boldsymbol{q}_{tt} = \boldsymbol{q}^{n+1} - 2\boldsymbol{q}^n + \boldsymbol{q}^{n-1}$. Note that the form given in Eq. (10) is only possible by using the linearization in Eq. (5) which allows the semi-implicit method to be written as a correction to the explicit method. This can be seen by taking $\delta = 0$ and equating Eqs. (12) and (11) which results in $q^{n+1} = q^{\text{explicit}}$. It should be mentioned that constructing the semi-implicit method as a correction to an explicit method as shown in Eq. (10) has been adopted universally by many, if not most, of the operational NWP centers including: the European Center for Medium-Range Weather Forecasting (see Ritchie et al. 1995), the National Center for Environmental Prediction, and the U. S. Navy.

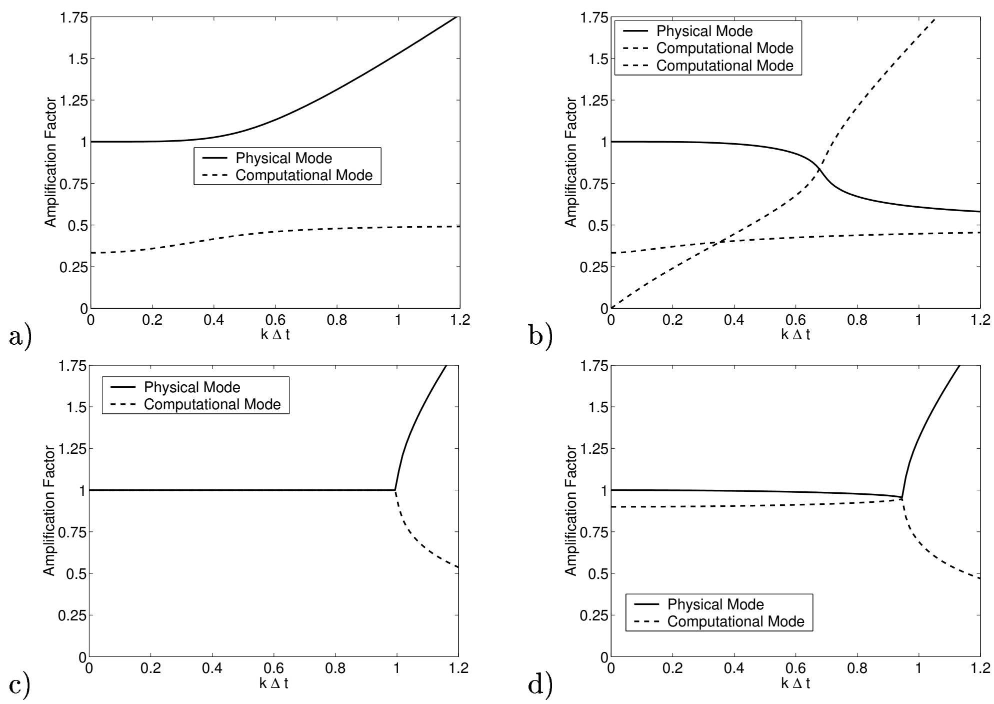

Figure 1. The stability of the explicit versions of a) BDF2A, b) BDF2B, c) LF2 with $\epsilon = 0$, and d) LF2 with $\epsilon = 0.05$.

In the following discussion we use the equation

$$\frac{\partial q}{\partial t} = ikq$$

to evaluate the stability properties of BDF2 and LF2. In Fig. 1 we show the stability region for the explicit formulations of BDF2 and LF2. Because BDF2 and LF2 both yield multiple numerical solutions to a first order equation this then means that only one solution is physical while the others are purely computational. The solid lines represent the physical solutions and the dashed lines are the computational modes. BDF2A clearly is inferior to the other two methods because it becomes unstable quite early near $k\Delta t = 0.15$ (Fig. 1a); however, the computational solution remains damped and thereby does not require the use of a filter. The physical solution of the BDF2B method is completely stable but becomes quite damped for increasing time step; however, one of the computational solutions eventually becomes unstable near $k\Delta t = 0.73$ (Fig. 1b). In contrast, since the original LF2 scheme is symplectic (i.e., a method which exactly conserves all Lagrangian integral invariants), it is completely undamped (Fig. 1c). This is true for both the physical and computational modes. For the physical mode this is a highly desirable property but not for the computational mode because it can become excited through nonlinear interactions with the physical mode and eventually become unstable (see Sanz-Serna 1985 and Aoyagi 1995). For this reason, in the geophysical fluid dynamics community LF2 is typically used in conjunction with a time filter (Asselin 1972) which damps the computational mode while selectively modifying the physical mode (Fig. 1d). This is by no means the only choice available for eradicating the computational mode. In the astrophysics community, which places much importance on symplecticness (which is ideal for Hamiltonian systems such as the first order HPE) the approach has been to introduce second order Runge-Kutta smoothers (see Aoyagi 1995 and New et al. 1998) which obviate the need for time-filters; however, in the present work we only consider LF2 with the Asselin time-filter.

Based on the above discussion it is safe to conclude that BDF2B and LF2 with $\epsilon = 0.05$ are the most stable schemes shown in Fig. 1. Let us now compare these two methods. Clearly, the ideal method would be one whereby the physical solution is undamped while the computational solution is damped. BDF2B and LF2 with $\epsilon = 0.05$ damp both the physical and computational solutions. The question we now try to answer is how much unwanted dissipation do these two methods introduce? Assuming that 1% numerical dissipation is an acceptable level (i.e., the amplification factor approaches 0.990) we find that BDF2B reaches this value at $k\Delta t = 0.38$ while LF2 with $\epsilon = 0.05$ at $k\Delta t = 0.58$. Thus BDF2B is much more dissipative than LF2 with $\epsilon = 0.05$. However, for $\epsilon = 0.1$ LF2 reaches this level at $k\Delta t = 0.41$ which is as dissipative as BDF2B. In the present work we use $\epsilon = 0.05$ for LF2 for most of the simulations but for very long time-integrations (such as the Held-Suarez test case) we use $\epsilon = 0.1$.

In Fig. 2 we show the stability region for the implicit formulations of BDF2 and LF2. For the fully implicit formulation ($\delta = 1$) BDF2A and BDF2B collapse to the classical BDF2 method found in numerical analysis text books (e.g., Gear 1971 page 217). The reason there is no computational mode for BDF2 in Fig. 2a is because the amplification factor of this mode remains below 0.32 and hence is not visible in this amplified plot. Clearly BDF2 and LF2 are unconditionally stable. However, this stability is achieved at the price of damping the physical solution. Once again, if we take 1% dissipation as an acceptable level we find that BDF2 reaches it at $k\Delta t = 0.51$ (Fig. 2a) while LF2 with $\epsilon = 0.05$ and $\vartheta = 0.5$ at $k\Delta t = 0.73$ (Fig. 2b). For $\epsilon = 0.1$ and $\vartheta = 0.5$ LF2 reaches this level at $k\Delta t = 0.46$ (Fig. 2c) which is now below BDF2. Changing the implicit weight of LF2 has even more drastic consequences. For example, for $\epsilon = 0.05$ and $\vartheta = 0.6$ LF2 reaches 1% dissipation at $k\Delta t = 0.22$ (Fig. 2d). In fact for this choice of ϵ and ϑ LF2 is more dissipative than BDF2 for the entire range of $k\Delta t$ values. Values of $\vartheta > 0.5$ are

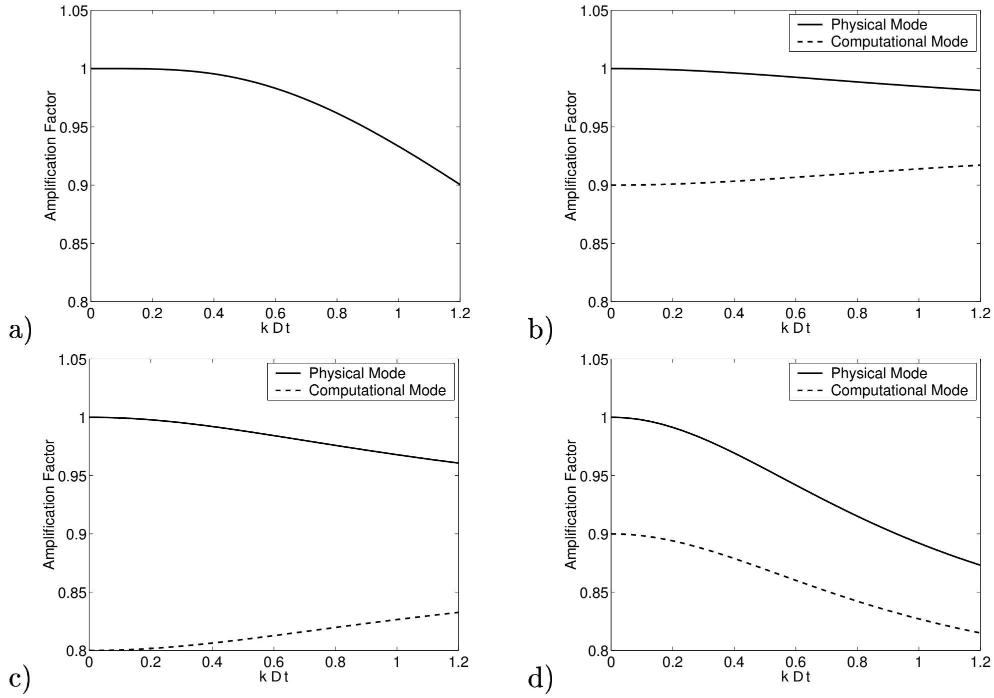

Figure 2. The stability of the implicit versions of a) BDF2, b) LF2 with $\epsilon = 0.05$ and $\vartheta = 0.5$, c) LF2 with $\epsilon = 0.1$ and $\vartheta = 0.5$, and d) LF2 with $\epsilon = 0.05$ and $\vartheta = 0.6$.

not uncommon in NWP models and in fact are often used to off-center semi-Lagrangian schemes in order to eliminate unwanted noise near steep topography (see Rivest and Robert 1994, Côté et al. 1995, and Davies et al. 2004). BDF2 has the advantage of being naturally off-centered and therefore able to avoid the spurious resonant response induced by orographic forcing.

In summary, LF2 with $\epsilon = 0$ and $\vartheta = 0.5$ is a better method than BDF2 because it is non-dissipative; however, as shown by Sanz-Serna (1985) LF2 can become unstable due to the nonlinear interaction between the physical and computational modes even when linear stability analysis shows otherwise. By choosing $\epsilon > 0$ and/or $\vartheta > 0.5$ LF2 becomes more stable but at the price of becoming dissipative and losing its second order accuracy. Thus if one is willing to accept a small amount of dissipation, have off-centering built into the time-integrator, but would like to retain second order accuracy then BDF2 is a reasonable choice. The stability analysis also shows that due to its larger explicit stability region LF2 admits a larger time step than BDF2, at least for the Rossby waves. However, as we shall see in Sec. 3 (b) , a larger time step does not necessarily translate into a faster model - at least not for pure dynamics (i.e., no diabatic forcing). The success of BDF2 for the shallow water equations using a spectral element semi-Lagrangian method (Giraldo et al. 2003), the incompressible Navier-Stokes equations using Eulerian time-integrators (Karniadakis et al. 1991), semi-Lagrangian methods (Xiu and Karniadakis 2001), and operator-integration-factor splitting methods (Maday et al. 1990) has been the main motivation for considering BDF2 for the semi-implicit version of

NSEAM. We now turn to the construction of the Helmholtz operator resulting from the generalized semi-implicit time-integration.

(b) *The 3D Pseudo-Helmholtz Operator*

After a standard application of the semi-implicit method, as outlined in Appendix A for the generalized 2nd order form given in Eq. (8), we obtain the following 3D pseudo-Helmholtz equation

$$\Phi_L - \lambda^2 V_{L,M} \left(M^{-1}\boldsymbol{D}^T\mathcal{P}M^{-1}\boldsymbol{D}\Phi\right)_M = \widehat{\Phi}_L - \lambda \boldsymbol{V}_{L,M} \left(M^{-1}\boldsymbol{D}^T\mathcal{P}\widehat{\boldsymbol{u}}\right)_M \tag{13}$$

for the variable Φ which is a linear combination of the potential temperature and surface pressure semi-implicit corrections (see Eq. (A.12)). In Eq. (13) M and $\boldsymbol{D}$ are the mass and differentiation matrices resulting from the spectral element discretization, $\mathcal{P}$ is the projection matrix which constrains the 3D Cartesian velocities to remain on the sigma levels, and V is the matrix containing the vertical contribution of the semi-implicit method (see Eq. (A.14)). At this point we have only used subscripts for the matrices corresponding to the vertical discretization. For completeness, below we include subscripts for the matrices corresponding to the horizontal discretization as well.

To simplify matters further let us define two additional horizontal matrices. Let

$$H^L_{I,J} = M^{-1}\boldsymbol{D}^T\mathcal{P}M^{-1}\boldsymbol{D} \tag{14}$$

represent the discrete pseudo-Laplacian operator and

$$\boldsymbol{H}^D_{I,J} = M^{-1}\boldsymbol{D}^T\mathcal{P}$$

the discrete divergence operator. Using these definitions along with some simplifications we can now write the Helmholtz operator, Eq. (13), in the following compact form

$$\left(I_{L,M}\otimes I_{I,J} - \lambda^2 (V_{L,M}\otimes H^L_{I,J})\right)\Phi_{J,M} = \widehat{\Phi}_{L,I} - \lambda(V_{L,M}\otimes \boldsymbol{H}^D_{I,J})\widehat{\boldsymbol{u}}_{J,M} \tag{15}$$

where $\otimes$ denotes the tensor product, $I_{L,M}$ and $I_{I,J}$ are identity matrices associated with the vertical and horizontal discretizations, respectively, and the range of the indices are $K, L, M = 1, ..., N_{lev}$ and $I, J = 1, ..., N_p$ where N_{lev} denotes the number of vertical layers and N_p the number of grid points in the horizontal direction (i.e., on each sigma layer). At this point, the Helmholtz operator of the semi-implicit formulation results in a fully 3D matrix. Below we describe the vertical mode decomposition which transforms the 3D HPE into a series of 2D shallow water equations which are then solved much more efficiently.

(c) *Vertical Mode Decomposition*

Note that the matrix H^L in Eq. (15) is completely independent from the vertical direction. However, the matrix V has a functional dependence on both the horizontal and vertical directions (see Appendix A). Thus, the optimum strategy is to apply a mode decomposition of the matrix V in the vertical direction. Upon completion of this decomposition the full 3D Helmholtz matrix problem will be converted into a series of N_{lev} 2D Helmholtz problems.

In order to perform this vertical decomposition we write the matrix V in the canonical form

$$V = R\Lambda R^{-1}$$

where Λ are the eigenvalues of V and R are the associated right eigenvectors which satisfy

$$VR = R\Lambda.$$

Note that this eigenvalue problem is not computationally expensive because V is an $N_{\text{lev}} \times N_{\text{lev}}$ matrix which only needs to be decomposed once. Upon obtaining this vertical mode decomposition we simply left-multiply Eq. (15) by R^{-1} to yield

$$\left(I_{I,J} - \lambda^2 \Lambda_L H^L_{I,J}\right) \left(R^{-1}_{L,M} \Phi_{M,J}\right)^T = \left[R^{-1}_{L,M} \left(\widehat{\Phi}_{M,I} - \lambda (V_{M,K} \otimes \boldsymbol{H}^D_{I,J}) \widehat{\boldsymbol{u}}_{J,K}\right)\right]^T . \tag{16}$$

Letting

$$\Phi^R_{J,L} = \left(R^{-1}_{L,M} \Phi_{M,J}\right)^T$$

we then obtain the following series of 2D Helmholtz problems

$$\left(I_{I,J} - \lambda^2 \Lambda_L H^L_{I,J}\right) \Phi^R_{J,L} = \left[R^{-1}_{L,M} \left(\widehat{\Phi}_{M,I} - \lambda (V_{M,K} \otimes \boldsymbol{H}^D_{I,J}) \widehat{\boldsymbol{u}}_{J,K}\right)\right]^T \tag{17}$$

for the variable Φ^R. Note that for each of the $L = 1, ..., N_{\text{lev}}$ modes Eq. (17) represents an individual 2D Helmholtz problem of size $N_p \times N_p$ which is in fact analogous to a shallow water model scaled by the eigenvalue Λ_L.

Figure 3a shows that the eigenvalues, Λ, of the six fastest vertical modes decrease exponentially. In fact, the third eigenvalue (Λ_2) has a value of 10^4 which its square root is already in the range of the highest horizontal wind velocities encountered in the atmosphere ($\sim$ 100 m/s). Thus the semi-implicit method is only required for the vertical modes which have gravity wave speeds beyond the highest horizontal wind velocities. For this reason we only solve Eq. (17) for the first four vertical modes in much the same way first proposed by Burridge (1975). Figure 3b shows the right eigenvectors, R, as a function of σ for the three fastest vertical modes. Let us now turn our attention to the solution of the 2D Helmholtz problems.

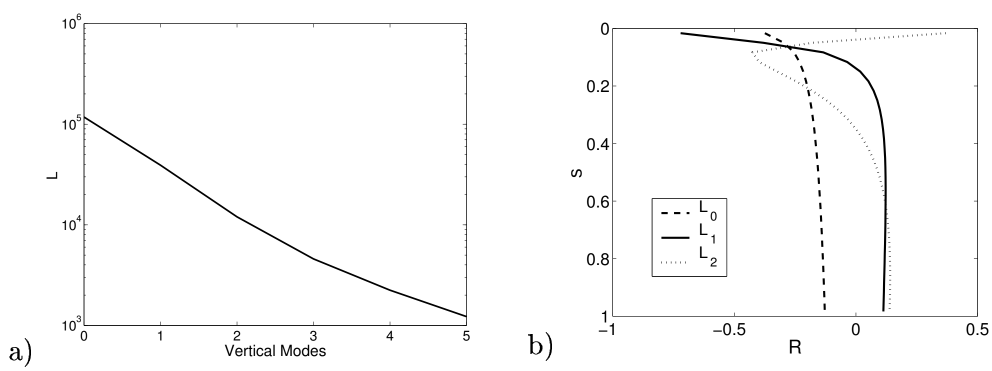

Figure 3. The a) eigenvalues, Λ, of the six fastest vertical modes and b) the right eigenvectors, R, of the three fastest vertical modes.

(*d*) *Solution of the 2D Pseudo-Helmholtz Operator*

From the vertical mode analysis it was determined that only the Helmholtz operator corresponding to the first four vertical modes need to be solved implicitly in time; the remainder can be computed explicitly. This coupling between explicit and implicit methods can be seamlessly included into the numerical approach by using the explicit solution, $\widehat{\Phi}$, as the initial guess for the implicit solution, Φ. This is why it is so beneficial to write the semi-implicit method as a correction to the explicit method. First, the explicit solution is obtained and then only for the first four vertical modes are the semi-implicit corrections included. The semi-implicit correction is obtained by solving Eq. (17) using the generalized minimum residual method (GMRES) with point Jacobi preconditioning, L_2 projection for the next iterate, and restarts every 10 time-steps (see Fischer et al. 1998). There are numerous other Krylov subspace methods to choose from but we have decided on GMRES based on previous experiences (see Giraldo et al. 2003). In addition, there are also more elaborate preconditioners and we shall report on our experiences with methods such as overlapping Schwarz (see Pavarino 2002) in future work.

The number of GMRES iteration required for convergence are dependent on the stopping criterion, ϵ_{stop}, and the size of $\lambda^2\Lambda$ in Eq. (17); the smaller ϵ_{stop} or the larger $\lambda^2\Lambda$ the more iterations required. The value of ϵ_{stop} is defined by the user and there are numerous strategies for choosing this value which are beyond the scope of the present work. It turns out that the value of $\lambda^2\Lambda$ is dependent on the eigenmode (more iterations are required for the external mode), the coefficients γ and ρ_{-1} of the semi-implicit method (see Eq. (A.6)), and of course the time-step. It is shown in Sec. 3 (*b*) that increasing the time-step by a given factor does not mean that an efficiency gain of this size will be achieved. Similarly, one time-integrator running with a larger time-step may not be more efficient than another time-integrator running with a smaller time-step. The overall efficiency of the model is determined by the scaling $\lambda^2\Lambda$ which affects the condition number of the resulting Helmholtz problem.

Upon obtaining the solution for Φ^R we then left-multiply by R to obtain Φ, that is,

$$\Phi = R\Phi_R.$$

Once Φ is obtained we then solve for $\boldsymbol{u}_{tt}$ via Eq. (A.15). With this value of $\boldsymbol{u}_{tt}$ known we can then solve for π_{tt} via Eq. (A.7) and θ_{tt} via Eq. (A.9). Finally, the prognostic variables are extracted from $\boldsymbol{q}_{tt}$ via Eq. (12). This then concludes the solution strategy for each time-step.

3. Parallel Implementation

(*a*) *Model Scalability*

One of the main advantages of using spectral element (SE) methods over spectral transform (ST) methods is that for an equivalent grid resolution the SE method allows the use of far more processors. Assuming that a 1D decomposition along latitude rings is the most efficient decomposition for ST models (as in the U. S. Navy's operational NWP model, NOGAPS), the maximum number of processors that a ST model can use is

$$N_{\text{proc}}^{\text{ST}} = N_{\text{lat}} \approx \frac{3}{2}T \tag{18}$$

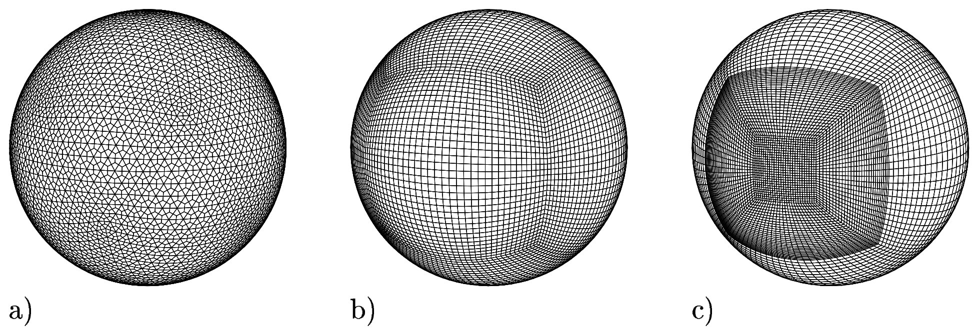

Figure 4. Three of the many possible grids that NSEAM can use. They include: a) icosahedral, b) hexahedral, and c) telescoping grids. The three grids have approximately the same number of grid points as the spectral grid $T239$ where the 49 high-order grid points inside each quadrilateral have been omitted for clarity.

where N_{lat} denotes the number of latitude rings and T the resolution of the spectral triangular truncation. In contrast, on a hexahedral grid (see Fig. 4b) with $N_p = 6(n_H N)^2 + 2$ grid points and $N_e = 6n_H^2$ elements (where n_H and N are the number of elements in each of the x, y directions on each of the six faces of the hexahedron and the polynomial order, respectively) the maximum number of processors that a SE model can use is

$$N_{\text{proc}}^{\text{SE}} = N_{\text{e}} \equiv \frac{6}{N^2} H^2 \tag{19}$$

where $H = n_H N$ represents the hexahedral horizontal resolution. In other words a SE model can use as many processors as there are elements. Thus for fixed N the number of processors allowed by a SE model increases quadratically with resolution, H, while only linearly for a ST model. At T239 (which is the current operational resolution used by NOGAPS) a ST model can use 360 processors. Using the approximation based on equivalent number of grid points

$$H \sim \frac{\sqrt{3}}{2}(T+1)$$

a spectral resolution of T239 translates to a hexahedral resolution of H208 which for simplicity we round-up to H216 (e.g., $n_H = 27$ $N = 8$, $n_H = 36$ $N = 6$, or $n_H = 54$ $N = 4$). At this resolution and assuming $n_H = 36$ $N = 6$ a SE model can use 7700 processors; a twenty-fold increase in the number of processors. Equation (19) shows that if we wish to further increase the number of processors of the SE model we simply increase n_H while decreasing N accordingly in order to maintain the horizontal resolution fixed. Therefore we could use $n_H = 54$ and $N = 4$ for a total of 17,000 processors; a forty-fold increase in the number of processors. However, decreasing N will impact the solution accuracy and the issue of efficiency versus accuracy must be carefully weighed. The point here is that the SE method offers this flexibility to increase either the accuracy or efficiency - a luxury shared by neither the spectral transform nor finite difference methods which are both built directly onto a specific grid geometry. Since SE methods are element-based they are built onto an element shape regardless of how the elements tile the sphere.

Figure 5 shows a comparison between the dynamical cores (i.e., no diabatic forcing) of NSEAM and NOGAPS on an IBM SP4 using single precision (all

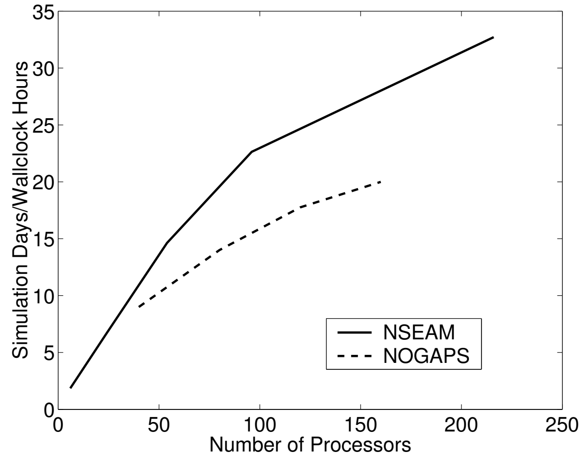

Figure 5. Simulation days per wallclock hours for the dynamical cores of NSEAM (T249 L30 using BDF2B) and NOGAPS (T239 L30) on an IBM SP4 using $\Delta t = 300$ seconds. (The performance data for NOGAPS is courtesy of Tim Hogan.)

other results from here on use double precision); NSEAM uses the semi-implicit method based on BDF2B. For NOGAPS we use the current operational resolution T239 L30 while for NSEAM we use T249 L30. At T239 L30 the maximum time step that NOGAPS can use is 300 seconds. At this time step using 96 processors NSEAM is 60% faster than NOGAPS; this gap between the two models will continue to widen with increased resolution in favor of NSEAM as predicted by Eqs. (53) and (54) in Giraldo and Rosmond (2004).

To see how NSEAM scales with increased resolution in Fig. 6 we show the simulation days per wallclock hours on 96 processors of an IBM SP4. At the resolution T580 L60 NSEAM can achieve 3 simulation days per wallclock hour on 96 processors. There are regions of the NSEAM model which can be improved to further increase its performance. Because NSEAM is currently only a research tool it has been designed with flexibility in mind. For example, the grids and the corresponding domain decomposition (DD) are generated automatically within the code. This consumes precious CPU time which could be avoided if the grid and its DD were generated off-line and then read in as input. This feature is currently being implemented into the next version of NSEAM which will have real terrain, diabatic forcing, and non-reflecting boundary conditions. We now turn our attention to the performance comparisons of the BDF2 and LF2 semi-implicit time-integrators.

(*b*) *Comparison of BDF2 and LF2 Semi-Implicit Time-Integrators*

Efficiency is arguably one of the most important criterion for determining whether a specific algorithm will be included in an operational NWP model. In Fig. 7 we show the performance of NSEAM using the BDF2B and LF2 semi-implicit time-integrators. The results for BDF2A are quite similar to those for BDF2B and so we shall refer to the BDF methods simply as BDF2. The resolution is T249 L30 with $\Delta t = 300$ seconds for BDF2 (solid line) and LF2 (dotted line) and $\Delta t = 200$ seconds for BDF2 (dashed line). The results in this figure show that the BDF2 semi-implicit formulation is more efficient than LF2 even when

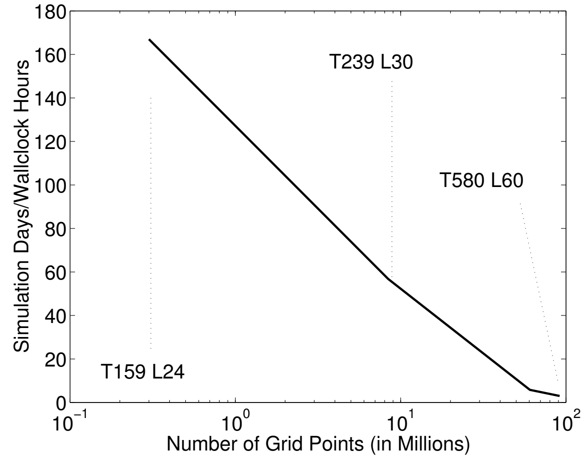

Figure 6. Simulation days per wallclock hours for the dynamical core of NSEAM (BDF2B) for various resolutions using 96 processors of an IBM SP4.

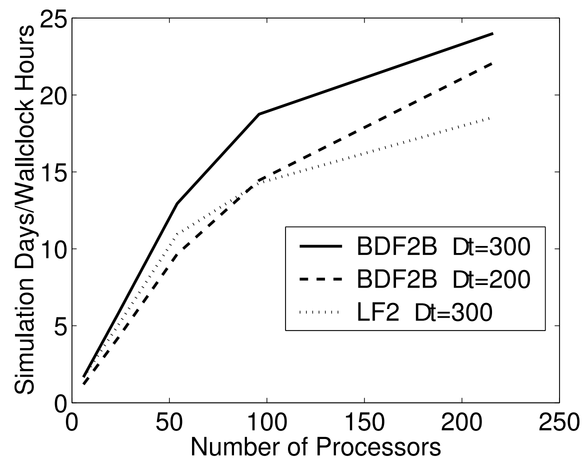

Figure 7. Efficiency of BDF2 and LF2 semi-implicit methods for NSEAM T249 L30.

BDF2 uses a time step 50 % smaller. The reason for this is quite simple: the 2D Helmholtz problem, given in Eq. (17), which must be solved at each time step only differs for the two methods by the parameter λ. For LF2 it is equal to $2\vartheta\Delta t$ and for BDF2 it is $\frac{2}{3}\Delta t$. Thus we can see that the matrix corresponding to BDF2 is more diagonally dominant than that for LF2 (since $\vartheta > \frac{1}{2}$), which gives BDF2 a smaller condition number than LF2. A smaller condition number means that the iterative solver will require fewer iterations for a given convergence tolerance; in Fig. 7, on average, BDF2 required 10 GMRES iterations per time step while LF2 required 21. This difference in GMRES iterations allows BDF2 to use a smaller time step and still perform as efficiently as LF2. Another interesting result is in the performance of BDF2. For this method using a time-step 50% smaller (dashed line) the decrease in performance was not significant especially at high processor counts. BDF2 with $\Delta t = 200$ seconds required fewer GMRES iterations than with

$\Delta t = 300$ seconds. Clearly, this has a dramatic effect on performance especially at high processor counts.

4. Results

The test cases used to validate NSEAM consist of five baroclinic tests. The difficulty with quantifying the error and/or accuracy of baroclinic models is that analytic solutions are difficult to obtain. Instead we view the following test cases as a means for qualitative comparisons to show that NSEAM gives similar results to existing models. In order to judge and compare the accuracy of NSEAM we plot normalized L_2 error norms defined as follows

$$||q||_{L_2} = \sqrt{\frac{\int_{\mathrm{A}} (q_{\mathrm{exact}} - q)^2 \, dA}{\int_{\mathrm{A}} q_{\mathrm{exact}}^2 dA}} \tag{20}$$

where q is the computed solution vector, q_{exact} is the exact solution, and A represents the surface area of the Earth.

(a) *Rossby-Haurwitz Wave Number 4*

In this test case (see Giraldo and Rosmond 2004) we track the propagation of Rossby-Haurwitz waves during a five day period. Surface pressure contours after a 5 day integration for NSEAM T185 L24 with $\Delta t = 300$ seconds for BDF2B and LF2 semi-implicit formulations are shown in Fig. 8. The results between the two methods are virtually indistinguishable; however, BDF2B required 33 % fewer iterations per time step to converge.

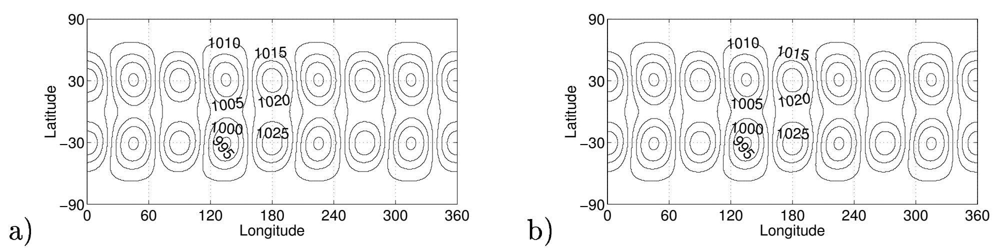

Figure 8. Rossby-Haurwitz Wave Number 4: The surface pressure contours in hPa for NSEAM T185 L24 using the a) BDF2B and b) LF2 semi-implicit time-integrators.

(b) *Polvani et al. Baroclinic Instability with Diffusion*

For this test case (see Polvani et al. 2004), the atmosphere is initially balanced and a perturbation is added to the flow which begins the motion of the atmosphere. This perturbation drives the atmosphere towards a singularity; however, in order to avoid this unpleasantness a diffusion operator is added to the momentum and thermodynamic equations with viscosity $7.0 \times 10^{-5}\ m^2/sec$. Figure 9 shows the surface temperature as a function of longitude and latitude at day 12 of the integration for the BDF2B and LF2 semi-implicit methods. The results are essentially identical between the two methods. In addition, these results compare extremely well with the Geophysical Fluid Dynamics Laboratory's (GFDL) spectral transform model (see Polvani et al. 2004). In fact, at this resolution NSEAM has converged to the correct solution regardless of which semi-implicit method is used. Further increases in either horizontal or vertical resolution show no discernible difference in the solution.

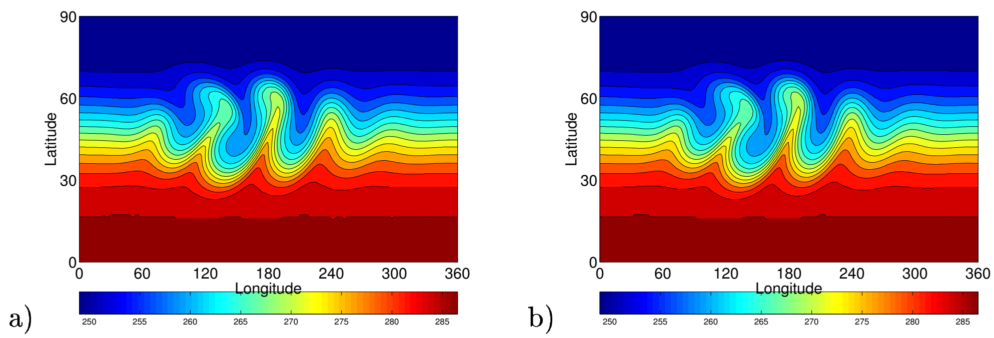

Figure 9. Polvani et al. Test Case: The surface temperature of NSEAM T74 L20 after 12 days using the a) BDF2B and b) LF2 semi-implicit time-integrators.

(*c*) *Jablonowski-Williamson Balanced Initial State*

For this test case (see Jablonowski and Williams 2002), the atmosphere is initially balanced and should remain so indefinitely. Figure 10 shows the normalized π L_2 error norm as a function of time for a 30 day period for NSEAM with the semi-implicit BDF2A (solid line), semi-implicit BDF2B (dashed-dotted line), semi-implicit LF2 (dotted line), and explicit LF2 (dashed line) methods for the resolution T185 L26; the explicit method used in this figure is in fact the model described in Giraldo and Rosmond (2004). Note that all the methods use different time steps with the explicit using the smallest (42 seconds) and the semi-implicit LF2 the largest (270 seconds). Even though the semi-implicit BDF2 and LF2 methods use time steps much larger than the explicit method the error norms are quite similar which confirms that the semi-implicit solutions are as accurate as the explicit one. It is surprising, however, that both semi-implicit BDF2 methods yield virtually indistinguishable results from the explicit method but that the semi-implicit LF2 method does not. For this case LF2 uses a time step 50 % larger than BDF2B but is only 210 seconds faster for the entire 30 day simulation. The advantages of the larger time step for LF2 are somewhat offset by requiring 50 % more iterations per time step than BDF2B.

(*d*) *Jablonowski-Williamson Baroclinic Instability*

This case is similar to the balanced initial state except that now a perturbation is added to the initial zonal velocity. This perturbation grows until a baroclinic instability develops near day nine. Figure 11 shows the minimum surface pressure, p_s, as a function of time for NSEAM (BDF2B) against various models including the NCAR ST model (Hack et al. 1992), the NASA Goddard FV model (Lin and Rood 1996), and the German Weather Service icosahedral FD model (Majewksi et al. 2002) which we denote as GME. For the grid point models, we use the definition of equivalent triangular truncation

$$T \sim \frac{\sqrt{2N_p}}{3}.$$

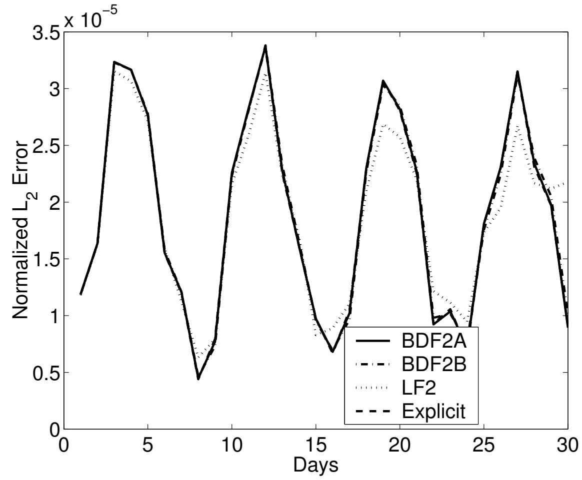

Figure 10. Jablonowski-Williamson Balanced Initial State: The normalized π L_2 error as a function of days for NSEAM T185 L26 using the semi-implicit BDF2A method with $\Delta t = 135$ seconds (solid line), semi-implicit BDF2B method with $\Delta t = 180$ seconds (dashed-dotted line), semi-implicit LF2 method with $\Delta t = 270$ seconds (dotted line), and the explicit LF2 method with $\Delta t = 42$ seconds (dashed line).

The results of this case are summarized as follows. Figure 11 shows that all four models are in complete agreement until day 8, at which point the two low-order models (NASA and GME) diverge from the NCAR and NSEAM models. The two low-order models, NASA and GME, show a similar pattern during the 14 day integration but do not match exactly. On the other hand the two high-order models, NCAR and NSEAM, behave almost identically throughout the entire 14 day simulation. Having established that NSEAM using BDF2B behaves similarly

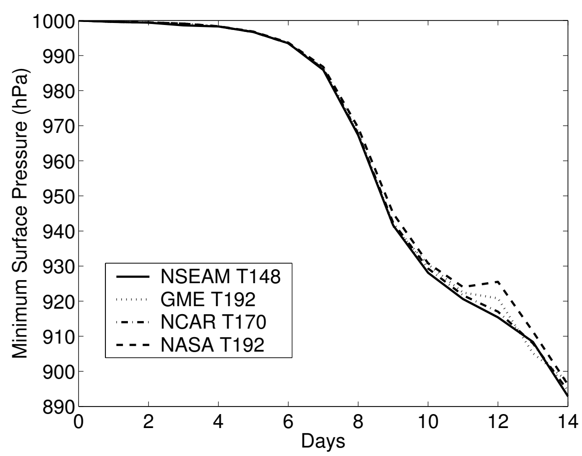

Figure 11. Jablonowski-Williamson Baroclinic Instability: The minimum surface pressure (hPa) as a function of days for the NASA (finite volume), GME (finite-difference), NCAR (spectral transform), and NSEAM (spectral element using BDF2B) models using 26 vertical levels. (The data for the last three models are courtesy of Christiane Jablonowski.)

to other well-established climate and NWP models we compare various semi-implicit time-integrators of NSEAM.

In Fig. 12 we plot the minimum surface pressure as a function of days for NSEAM T185 L26 using the semi-implicit BDF2A (solid line), semi-implicit BDF2B (dotted line), semi-implicit LF2 (dashed-dotted line), and the explicit LF2 (dashed line). Even though all the models use different time steps they all agree rather well confirming that the semi-implicit time-integrators are as accurate as the explicit method.

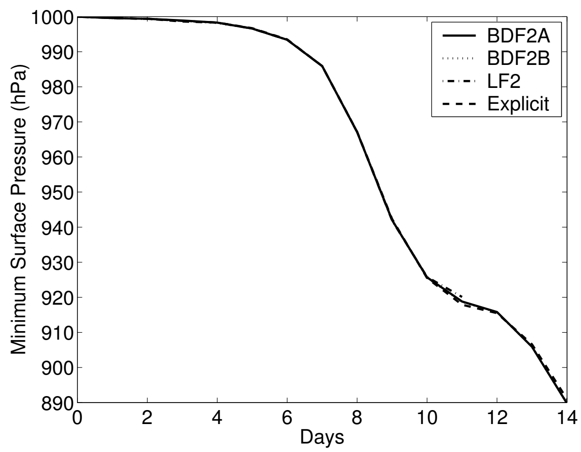

Figure 12. Jablonowski-Williamson Baroclinic Instability: The minimum surface pressure (hPa) as a function of days for NSEAM T185 L26 using the semi-implicit BDF2A with $\Delta t = 135$ seconds (solid line), semi-implicit BDF2B method with $\Delta t = 180$ seconds (dotted line), semi-implicit LF2 method with $\Delta t = 270$ seconds (dashed-dotted line), and the explicit LF2 method with $\Delta t = 42$ seconds (dashed line).

(*e*) *Held-Suarez Mean Planetary Climate*

For this test case (see Held and Suarez 1994), the atmosphere is initially at rest and a perturbation is added to the flow which begins the motion of the atmosphere. A forcing function mimicking the radiation of the sun near the Equator drives the model towards a realistic mean planetary climate. NSEAM was run for 1200 days with samples taken every 4 days beginning from day 200. The sample files from day 200 to day 1200 are then averaged to obtain a temporal mean. Figure 13 shows the time and zonally averaged zonal velocity for the semi-implicit BDF2B (left panel) and LF2 (right panel) as a function of latitude and vertical coordinate. Both models yield identical mid-latitude jets in the upper atmosphere which agree with the results obtained with the explicit model in Giraldo and Rosmond (2004) and the spectral transform model in Held and Suarez (1994).

5. Conclusion

The Naval Research Laboratory's spectral element atmospheric model (NSEAM) for scalable computer architectures was presented. NSEAM is based on a Cartesian formulation of the equations, the spectral element (SE) method in space, and a general second order semi-implicit method in time. Specifically, semi-implicit methods based on backward difference formulas (BDF2) and leapfrog (LF2) were compared. The stability analysis showed that for reasonable values of the Asselin time-filter ϵ LF2 is less dissipative than BDF2 as long as the implicit weight ϑ is

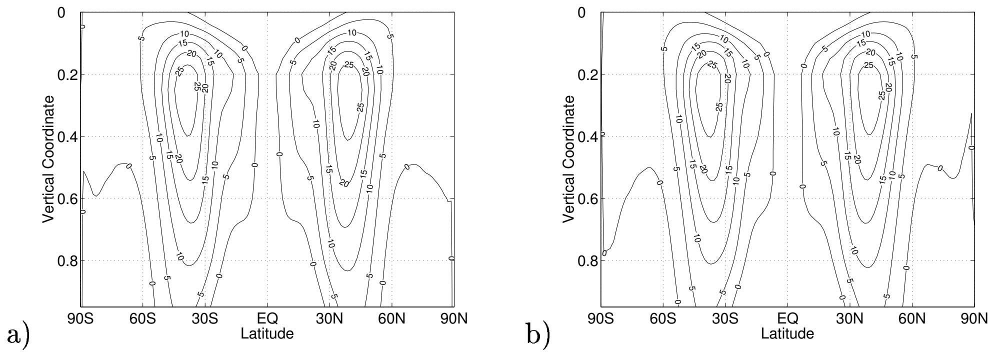

Figure 13. Held-Suarez Mean Planetary Climate: The time and zonally averaged zonal velocity of the semi-implicit a) BDF2B and b) LF2 methods as function of latitude and vertical coordinate for the resolution T74 L20.

equal to 0.5; if $\vartheta > 0.5$ then LF2 becomes more dissipative than BDF2. For pure dynamics BDF2 was shown to be as accurate as LF2 while being more efficient even though it required a smaller time-step. This difference in efficiency is due to the resulting Helmholtz matrix having a smaller condition number for BDF2 than LF2 which then translates into fewer iterations per time step to converge.

We showed that regardless of time-integrator NSEAM gives similar results to well-established climate and weather prediction models while scaling quite efficiently on distributed-memory computers. At a resolution of T249 L30 using 96 processors, NSEAM was shown to be 60% faster than a spectral transform model and this gap will continue to grow in favor of NSEAM as the horizontal resolution and the number of processors are increased. Because SE models are constructed completely around basis functions this offers attractive flexibilities not shared by other methods. The shape, order, and characteristics of the grid, polynomials, and model can be altered merely by changing the basis functions. This flexibility allows an SE model to adapt to the changing needs in science and computing throughout its lifetime. Finally, the advantage of using Cartesian coordinates with the SE method is that the model becomes completely independent from the grid. This means that any type of grid can be used with NSEAM. While we have only shown results on hexahedral grids, the extension to icosahedral, telescoping, and adaptive grids is immediately obvious. Various improvements in the accuracy, efficiency, and flexibility of NSEAM are currently underway and the results of this research will be reported in the future.

Acknowledgements

This work was supported by the Office of Naval Research through program element PE-0602435N. All computer simulations were performed on the IBM SP4 at the Department of Defense NAVO Major Shared Resource Center. I would like to thank Craig Bishop and João Teixeira for their helpful comments on an early draft of the manuscript. I also would like to thank Terry Davies and Lorenzo Polvani for allowing me access to their unpublished works. Finally, I would like to thank Andrew Staniforth, Clive Temperton, and the anonymous reviewer for their detailed review which helped to improved the manuscript.

Appendix A

Semi-Implicit Formulation

(a) Linearization of the Nonlinear Gravity Wave Terms S^G

The terms in Eq. (4) are linearized in the following manner. First, π and θ are linearized about the reference states $\pi^* = 1000$ hPa and

$$\theta^* = \frac{T^*}{P\left(\pi^*\right)}$$

with $T^* = 300°$ Kelvin (see Temperton et al. 2001). With these reference states defined we can now proceed with the linearization of the terms in Eq. (4).

(i) *Surface Pressure and Thermodynamic Equations*

Based on the previously defined reference states, the surface pressure and thermodynamic equations are linearized in the following simple form

$$S^{\mathrm{G}}(\pi) = -\pi^* \left(\boldsymbol{\nabla} \cdot \boldsymbol{u} + \frac{\partial \dot{\sigma}}{\partial \sigma} \right) \quad \text{and} \quad S^{\mathrm{G}}(\theta) = -\dot{\sigma} \frac{\partial \theta^*}{\partial \sigma},$$

respectively.

(ii) *Momentum Equation*

In order to linearize the pressure gradient the geopotential height itself must be linearized. Beginning with the finite differenced equation for the geopotential height

$$\phi_L - \phi_{L+1} = c_p \theta_L \left(P_{L+1/2} - P_L \right) + c_p \theta_{L+1} \left(P_{L+1} - P_{L+1/2} \right),$$

where $L = 1, ..., N_{lev}$ with N_{lev} representing the number of vertical levels, we can then take a Taylor series expansion about the reference states π^* and θ^* yielding

$$\begin{aligned} \phi_L - \phi_{L+1} &= c_p \theta_L \left(P^*_{L+1/2} - P^*_L \right) + c_p \theta_{L+1} \left(P^*_{L+1} - P^*_{L+1/2} \right) \\ &+ c_p \theta^*_L \left(\frac{\partial P^*_{L+1/2}}{\partial \pi} - \frac{\partial P^*_L}{\partial \pi} \right) (\pi - \pi^*) + c_p \theta^*_{L+1} \left(\frac{\partial P^*_{L+1}}{\partial \pi} - \frac{\partial P^*_{L+1/2}}{\partial \pi} \right) (\pi - \pi^*) . \end{aligned} \tag{A.1}$$

Equation $(A.1)$ can be written in matrix form as

$$A_{L,K} \phi_K = B_{L,K} \theta_K + C_L (\pi - \pi^*)$$

where

$$A_{L,K} = \begin{cases} 1 & \text{if} \quad K = L \\ -1 & \text{if} \quad K = L + 1 \\ 0 & \text{if} \quad K < L \text{ or } K > L + 1 \end{cases}$$

$$B_{L,K} = \begin{cases} c_p \left(P^*_{L+1/2} - P^*_L \right) & \text{if} \quad K = L \\ c_p \left(P^*_{L+1} - P^*_{L+1/2} \right) & \text{if} \quad K = L + 1 \\ 0 & \text{if} \quad K < L \text{ or } K > L + 1 \end{cases}$$

$$C_L = \begin{cases} c_p \theta^*_L \left(\frac{\partial P^*_{L+1/2}}{\partial \pi} - \frac{\partial P^*_L}{\partial \pi} \right) + c_p \theta^*_{L+1} \left(\frac{\partial P^*_{L+1}}{\partial \pi} - \frac{\partial P^*_{L+1/2}}{\partial \pi} \right) & \text{if} \quad L < N_{\mathrm{lev}} \\ c_p \theta^*_L \left(\frac{\partial P^*_{L+1/2}}{\partial \pi} - \frac{\partial P^*_L}{\partial \pi} \right) & \text{if} \quad L = N_{\mathrm{lev}} \end{cases}$$

and the range of the indices are K, L, $M = 1, ..., N_{lev}$. This results in the following geopotential gradient

$$\nabla \phi_L = \nabla \left(A^{-1}_{L,K} B_{K,M} \theta_M + A^{-1}_{L,K} C_K (\pi - \pi^*) \right)$$

where

$$A^{-1}_{L,K} = \begin{cases} 1 & \text{if} \quad K \geq L \\ 0 & \text{if} \quad K < L \end{cases} .$$

Linearizing the gradient due to surface pressure yields

$$\left(c_p \theta_L \frac{\partial P_L}{\partial \pi} \nabla \pi \right)^{n+1} = \left(c_p \theta^*_L \frac{\partial P^*_L}{\partial \pi} \nabla \pi^{n+1} \right) .$$

To simplify the description of the semi-implicit method let us create the following definitions: let

$$E_{L,M} = A^{-1}_{L,K} B_{K,M} \text{ and } F_L = A^{-1}_{L,K} C_K + c_p \theta^*_L \frac{\partial P^*_L}{\partial \pi},$$

which allows the gradient of geopotential and surface pressure to be written as

$$\nabla \phi_L + c_p \theta_L \frac{\partial P_L}{\partial \pi} \nabla \pi = \nabla \left(E_{L,K} \theta_K + F_L \pi \right) .$$

We are now ready to construct the linear operator $L^G(\boldsymbol{q})$ of Eq. (5).

(b) *Linear Operator and Implicit Correction*

With the linearizations described above and dropping the subscripts, the linear operator L^G in Eq. (5) can be written as follows

$$L^G(\boldsymbol{q}) = - \begin{pmatrix} \pi^* \left(\nabla \cdot \boldsymbol{u} + \frac{\partial \dot{\sigma}}{\partial \sigma} \right) \\ \nabla \left(E\theta + F\pi \right) \\ \dot{\sigma} \frac{\partial \theta^*}{\partial \sigma} \end{pmatrix} . \tag{A.2}$$

Using this linearization we can now write the equations in terms of the semi-implicit correction as follows

$$\pi_{tt} = \widehat{\pi} - \lambda \pi^* \left(\nabla \cdot \boldsymbol{u}_{tt} + \frac{\partial}{\partial \sigma} \left(\dot{\sigma}_{tt} \right) \right) \tag{A.3}$$

$$\boldsymbol{u}_{tt} = \widehat{\boldsymbol{u}} - \lambda \nabla \left(E\theta_{tt} + F\pi_{tt} \right) - \mu \boldsymbol{x} \tag{A.4}$$

$$\theta_{tt} = \widehat{\theta} - \lambda \left(\dot{\sigma}_{tt} \frac{\partial \theta^*}{\partial \sigma} \right) \tag{A.5}$$

where

$$\lambda = \delta \gamma \Delta t \rho_{-1} \tag{A.6}$$

and $\widehat{\boldsymbol{q}}$ and $\boldsymbol{q}_{tt}$ are defined in Eqs. (11) and (12).

(c) Construction of the 3D Pseudo-Helmholtz Operator

Replacing the continuous spatial operators in Eqs. (A.3), (A.4), and (A.5) by their discrete spectral element counterparts results in

$$\pi_{tt} = \widehat{\pi} - \frac{\lambda \pi^*}{\sigma_S - \sigma_T} \sum_{K=1}^{N_{\text{lev}}} M^{-1} \boldsymbol{D}^T \boldsymbol{u}_{tt\,K} \Delta\sigma_K \tag{A.7}$$

$$\boldsymbol{u}_{tt\,L} = \mathcal{P}\widehat{\boldsymbol{u}}_L - \lambda \mathcal{P} M^{-1} \boldsymbol{D} \left(E_{L,K}\theta_{tt\,K} + F_L \pi_{tt} \right) \tag{A.8}$$

$$\theta_{tt\,L} = \widehat{\theta}_L - \lambda S_{L,K} \left(M^{-1} \boldsymbol{D}^T \boldsymbol{u}_{tt} \right)_K \tag{A.9}$$

where

$$S_{L,K} = \Theta_L \Delta_{N_{\text{lev}},K} - \Theta_{L+1/2}\Delta_{L,K} - \Theta_{L-1/2}\Delta_{L-1,K}, \tag{A.10}$$

$$\Theta_L = \Theta_{L+1/2}\frac{\sigma_{L+1/2} - \sigma_T}{\sigma_S - \sigma_T} + \Theta_{L-1/2}\frac{\sigma_{L-1/2} - \sigma_T}{\sigma_S - \sigma_T},$$

$$\Theta_{L+1/2} = \frac{\theta^*_{L+1/2} - \theta^*_L}{\sigma_{L+1/2} - \sigma_L}, \quad \Theta_{L-1/2} = \frac{\theta^*_L - \theta^*_{L-1/2}}{\sigma_L - \sigma_{L-1/2}},$$

$$\Delta_{L,K} = \begin{cases} \Delta\sigma_K & \text{if} \quad K \le L \\ 0 & \text{if} \quad K > L \end{cases}, \quad \Delta\sigma_K = \sigma_{K+1/2} - \sigma_{K-1/2},$$

and M is the mass matrix, $\boldsymbol{D}$ is the differentiation matrix, and $\mathcal{P}$ is the projection matrix which constrains the Cartesian velocity field to remain tangential to a sigma surface (for details on the SE discretization see Giraldo and Rosmond 2004). To simplify the description of the semi-implicit formulation we only include subscripts for the matrices involving the vertical discretization because these are the terms to which the vertical mode decomposition is applied.

Multiplying Eq. (A.9) by E and Eq. (A.7) by F and adding gives

$$\Phi_L = \widehat{\Phi}_L - \lambda E_{L,K} S_{K,M} \left(M^{-1}\boldsymbol{D}^T \boldsymbol{u}_{tt}\right)_M - \lambda F_L \frac{\pi^*}{\sigma_S - \sigma_T} \Delta_{N_{\text{lev}},M} \left(M^{-1}\boldsymbol{D}^T \boldsymbol{u}_{tt}\right)_M \tag{A.11}$$

where

$$\Phi_L = E_{L,K}\theta_{tt\,K} + F_L \pi_{tt}, \tag{A.12}$$

$$\widehat{\Phi}_L = E_{L,K}\widehat{\theta}_K + F_L \widehat{\pi},$$

and σ_S and σ_T are the sigma values at the surface and top of the model. Let us next factor $M^{-1}\boldsymbol{D}^T\boldsymbol{u}_{tt}$ from Eq. (A.11) and rewrite the equation as

$$\Phi_L = \widehat{\Phi}_L - \lambda V_{L,M} \left(M^{-1}\boldsymbol{D}^T \boldsymbol{u}_{tt}\right)_M \tag{A.13}$$

where

$$V_{L,M} = E_{L,K}S_{K,M} + F_L \frac{\pi^*}{\sigma_S - \sigma_T} \Delta_{N_{\text{lev}},M} \tag{A.14}$$

is the matrix containing the vertical contribution to the semi-implicit formulation. Note that Eq. (A.8) can now be written as

$$\boldsymbol{u}_{tt\,L} = \mathcal{P}\left(\widehat{\boldsymbol{u}}_L - \lambda M^{-1}\boldsymbol{D}\Phi_L \right). \tag{A.15}$$

Substituting Eq. (A.15) into Eq. (A.13) yields

$$\Phi_L - \lambda^2 V_{L,M} \left(M^{-1}\boldsymbol{D}^T \mathcal{P} M^{-1}\boldsymbol{D}\Phi \right)_M = \widehat{\Phi}_L - \lambda V_{L,M} \left(M^{-1}\boldsymbol{D}^T \mathcal{P}\widehat{\boldsymbol{u}} \right)_M .$$

References

Aoyagi, A. 1995 Nonlinear leapfrog instability for Fornberg's pattern. *J. Comput. Phys.* **120** 316-322

Asselin, R. 1972 Frequency filter for time integrations. *Mon. Weather Rev.* **100** 487-490

Boyd, J. P. 2004 Prolate spheroidal wavefunctions as an alternative to Chebyshev or Legendre polynomials for spectral elements and pseudospectral algorithms. *J. Comput. Phys.* **199** 688-716

Burridge, D. M. 1975 A split semi-implicit reformulation of the Bushby-Timpson 10-level model. *Q. J. R. Meteorol. Soc.* **101** 777-792

Côté, J. 1988 A Lagrange multiplier approach for the metric terms of semi-Lagrangian models on the sphere. *Q. J. R. Meteorol. Soc.* **114**, 1347-1352

Côté, J., Gravel, S., and Staniforth, A. 1995 A generalized family of schemes that eliminate the spurious resonant response of semi-Lagrangian schemes to orographic forcing. *Mon. Weather Rev.* **123**, 3605-3613

Côté, J. C., Gravel, S., Methot, A., Patoine, A., Roch, M., and Staniforth, A. 1998 The operational CMC-MRB global environmental multiscale (GEM) model. Part I: design considerations and formulation. *Mon. Weather Rev.* **126**, 1373-1395

Davies, T., Cullen, M. J. P, Malcolm, A. J., Mawson, M., H., Staniforth, A., White, A. A., and Wood, N. 2004 A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meterol. Soc.* in press

Fischer, P. F. 1998 Projection techniques for iterative solutions of $\boldsymbol{Ax} = \boldsymbol{b}$ with successive righ-hand sides. *Computer Methods in Applied Mechanics and Engineering* **163**, 193-204

Fournier, .A, Bunge, H. P., Hollerbach, R., and Vilotte, J. P. 2004 Application of the spectral element method to the axisymmetric Navier-Stokes equations. *Geophys. J. Int.* **156**, 682-700

Gear, C. W. 1971 *Numerical initial value problems in ordinary differential equations.* Prentice-Hall, Englewood Cliffs, New Jersey

Giraldo, F. X., Hesthaven, J. S., and Warburton, T. 2002 Nodal high-order discontinuous Galerkin methods for the spherical shallow water equations. *J. Comput. Phys.* **181**, 499-525

Giraldo, F. X., Perot, J. B., and Fischer, P. F. 2003 A spectral element semi-Lagrangian (SESL) method for the spherical shallow water equations. *J. Comput. Phys.* **190**, 623-650

Giraldo, F. X., and Rosmond, T. E. 2004 A scalable spectral element Eulerian atmospheric model (SEE-AM) for NWP: dynamical core tests. *Mon. Weather Rev.* **132**, 133-153

Giraldo, F. X., and Warburton, T. 2004 A nodal triangle-based spectral element method for the shallow water equations on the sphere. *J. Comput. Phys.* in press

Hack, J. J., Boville, B. A., Briegleb, B. P., Kiehl, J. T., Rasch, P. J., and Williamson, D. L. 1992 Description of the NCAR community climate model (CCM2). *NCAR Technical Note NCAR/TN-382+STR*, National Center for Atmospheric Research, Climate Modeling Section, P.O. Box 3000, Boulder CO. 80307

Hogan, T. F., and Rosmond, T. E 1991 The description of the Navy Global Operational Prediction System's spectral forecast model. *Mon. Weather Rev.* **119**, 1786-1815

Hulsten, M. A. 1996 Stability of the implicit/explicit extension of the stiffly-stable schemes of Gear. *unpublished*

Iskandarani, M., Haidvogel, D. B., Levin, J. C., Curchitser E., and Edwards C. A. 2002 Multiscale geophysical modeling using the spectral element method. *Computing in Science and Engineering* **4**, 42-48

Jablonowski, C. and Williamson, D. L. 2002 Baroclinic instability test with two jets in the midlatitudes . Abstracts, *The 2002 Workshop on the Solution of Partial Differential Equations on the Sphere*, Toronto, Canada, 10-10

Karniadakis, G. E., Israeli, M., and Orszag, S. A. 1991 High-order splitting methods for the incompressible Navier-Stokes equations. *J. Comput. Phys.* **97**, 414-443

Knoll, D. A., and Keyes, D. E. 2004 Jacobian-free Newton-Krylov methods: a survey of approaches and applications. *J. Comput. Phys.* **193**, 357-397

Komatitsch, D., and Tromp, J. 1999 Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophys. J. Int.* **139**, 806-822

Kwizak, M., and Robert, A. J. 1971 A semi-implicit scheme for grid point atmospheric models of the primitive equations. *Mon. Weather Rev.* **99**, 32-36

Lin, S. J., and Rood, R. B. 1996 Multidimensional flux-form semi-Lagrangian transport schemes. *Mon. Weather Rev.* **124**, 2046-2070

Maday, Y., Patera, A. T., and Ronquist, E. M. 1990 An operator-integration-factor splitting method for time-dependent problems: application to incompressible fluid flow. *J. Scient. Computing* **5**, 263-292

Majewski, D., Liermann, D., Prohl, P., Ritter, R., Buchhold, M., Hanisch, T., Paul, G., Wergen, W., and Baumgardner, J. 2002 The operational global icosahedral-hexahedral gridpoint model GME: description and high-resolution tests. *Mon. Weather Rev.* **130**, 319-338

New, K. C. B., Watt, K., Misner, C. W., and Centrella, J. M. 1998 Stable 3-level leapfrog integration in numerical relativity. *Physical Review D* **58**, 064022, 1-14

Pavarino, L. F. 2002 Overlapping Schwarz preconditioners for spectral element discretizations of convection-diffusion problems. *International Journal for Numerical Methods in Engineering* **53**, 1005-1023

Polvani, L. M., Scott, R. K., and Thomas, S. J. 2004 Numerically converged solutions of the global primitive equations for testing the dynamical core of atmospheric GCMs. *Monthly Weather Review* in press

Randall, D. A., Ringler, T. D., and Heikes, R. P. 2002 Climate modeling with spherical geodesic grids. *Computing in Science and Engineering* **4**, 32-41

Ritchie, H., Temperton, C., Simmons, A., Hortal, M., Davies, T., Dent, D., and Hamrud, M. 1995 Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Monthly Weather Review* **123**, 489-514

Shen, J., and Wang, W. 1999 A fast and accurate numerical scheme for the primitive equations of the atmosphere. *SIAM J. Numer. Anal.* **36**, 719-737

Temperton, C., Hortal, M., and Simmons, A. 2001 A two-time-level semi-Lagrangian global spectral model. *Q. J. R. Meterol. Soc.* **127**, 111-127

Thomas, S. J., Loft, R. D., and Dennis, J. M. 2002 Parallel implementation issues: Global versus local methods. *Computing in Science and Engineering* **4**, 26-31

Trémolet, Y. and Sela, J. 1999 Parallelization of the NCEP global spectral model. in *Towards Teracomputing*, Proceedings of the 8th ECMWF Workshop on the Use of Parallel Processors in Meteorology, ed. W. Zwieflhofer and N. Kreitz, World Scientific, 109-116

Xiu, D., and Karniadakis, G. E. 2001 A semi-Lagrangian high-order method for the Navier-Stokes equations. *J. Comput. Phys.* **172**, 658-684